COMP 135 Introduction to Machine Learning

First day of class

Spring 2019

https://www.cs.tufts.edu/comp/135/2019s/

Many slides attributable to: Emily Fox (UW), Finale Doshi-Velez (Harvard), Erik Sudderth (UCI), & Liping Liu (Tufts)

Why Machine Learning?



Image Credit: Emily Fox

Artificial Intelligence (AI)

• Study of "intelligent systems", with many parts: logic, planning, search, probabilistic reasoning, **learning from experience**, interacting with other agents, etc.



Machine Learning (ML)

- Study of algorithms that **learn from experience/data** to perform a task
- Task output: a *prediction* or a *decision*





The Machine Learning Process



Why take this course?

Why take this course?

Our goal is to prepare students to effectively apply machine learning methods to problems that might arise in "the real world" -- in industry, medicine, education, and beyond.

Gain skills and understanding for a future as:

- Developer using ML "out-of-the-box"
- ML methods researcher

Why take this course?

Students will be able to:

- Think systematically
 - Compare/contrast each method's strengths & limitations
- Deploy rapidly
 - Hands-on experience with open-source libraries
 - 襣 python

- Evaluate carefully
 - Design experiments with task-appropriate performance metrics and strong baselines
 - Report uncertainty in performance

What skills will we learn?



What will we learn?



Task: Regression



is a continuous variable e.g. sales in \$\$



Regression Example: Uber

Supervised Learning

regression

Unsupervised Learning

Predictions of travel time, price, supply, demand



CEDERAL STORES.

SET PICKUP LOCATION

Reinforcement Learning

Regression Example: Uber



Regression Example: Uber



(Keith Chen)

Task: Classification



Classification Example: Swype

Predict words from keyboard trajectories



What will we learn?



What will we learn?



Task: Clustering

Supervised Learning

Unsupervised Learning

clustering

Reinforcement Learning



Clustering Example: News

Top Stories



Intensive manhunt underway after daring jail escape in California

Fox News - 37 minutes ago 🛛 🚱 💟 🔣

An intensive manhunt was underway Monday for three inmates who pulled a "Shawshank"-style escape through a hole in their California jail cell -- and, who may have ties to notorious Vietnamese street gangs and Iran.

Manhunt Expands For 'Dangerous' Trio After Daring Jailbreak NBCNews.com

Orange County manhunt: Officials suggest violent jail escapees could

Related California »

23

\$

2

Fax News

See realtime coverage

be hiding nearby Los Angeles Times

In Depth: Authorities struggling to piece together daring jail escape Washington Post











Huffington Post

See realtime coverage

HUFFPOLLSTER: Trump And Clinton Lead, But Iowa Polling Remains Volatile With A Week To Go

Huffington Post - 5 hours ago 🛛 🔂 🛃 🖂

Donald Trump has regained the lead in Iowa but things can still change. On the Democratic side, young voters could tip the caucus toward Bernie Sanders, but only if they turn out.

Here's Bernie Sanders's best closing argument against Hillary Clinton In Jowa Washington Post Bernie Sanders' One Answer on How He Would Get Anything Done ABC News

Related Hillary Rodham Clinton » Bernie Sanders »

Opinion: Democratic lowa Forum: How to Watch the Live Stream Online Daily Beast Wikipedia: Statewide opinion polling for the Democratic Party presidential primaries, 2016



Task: Embedding



Example: Genes vs. geography



Task: Recommendation



Recommendation Example





RL example: Pancake robot



Peter Kormushev, Imperial College

What will we learn this semester?

Supervised Learning ~ 10 weeks 6 homeworks 3 projects If I want more?

Take COMP 136

Take COMP 150 – Deep Learning

Unsupervised Learning

~ 2 weeks 1.5 homeworks

Take COMP 136

Reinforcement Learning

~ 1 week 0.5 homeworks Take COMP 150 – RL (Prof. Jivko Sinapov)

What we won't cover

- Active learning
- Transfer learning
- Semi-supervised learning
- Learning theory
- lots more

Who is teaching?

- Instructor: Prof. Mike Hughes
- TA Staff
 - Mike Pietras
 - Rui Chen
 - Minh Nguyen
 - Duc Nguyen
 - Wayne Tang

Problem: When will ICU patient **need intervention**?

Ghassemi, Wu, Hughes, et al. AMIA CRI 2017

Interventions:

- mechanical ventilation
- blood pressure drugs

Early prediction helps: prepare patient plan staffing try less aggressive options early



Cohort from MIMIC-III dataset mimic.physionet.org

36,050 patients

(Johnson et al. Sci. Data 2016)

- from Beth-Israel Deaconess in Boston 2001-2012
- kept all adults with record within 6-360 hours

Intervention	Training Num Positive	Training Num Control	Heldout Num Positive	Heldout Num Control
Vasopressor	6987	21865	1737	5461
Red blood cell transfusion	19171	9681	4776	2422
Fresh frozen plasma transfusion	2759	26093	620	6578
Platelet transfusion	27818	1034	6944	254
Mechanical Ventilation	13710	15142	3393	3805

Observed data

7 nurse-validated vital signs (hourly)
heart rate, blood pressure, temp., SpO2, ...
11 lab measurements (much less than hourly)
hematocrit, lactate, ...



each channel standardized to mean=0, var=1 with carry-and-hold for missing data

Task: predict need **in advance**



Vasopressor prediction : 1 hr ahead



Key Stakeholder Questions

- How should we fill in missing values?
- How to deal with imbalance data?
 - most patients never get drug X
- How to deal with imbalanced mistake costs?
 - remove ventilator too early ends life, too late costs \$\$
- How uncertain are predictions?
- How will this generalize to a new hospital?

We can start to answer many of these in COMP 135

Logistics

- Course website (take a tour)
 - <u>https://www.cs.tufts.edu/comp/135/2019s/</u>
- Discussions on Piazza
- Lectures every Mon & Wed
- Recitation Sessions with TAs every Mon
- Deliverables: 2 exams, 8 homeworks, 3 projects
- Collaboration Policy

Let's Get Started!

- Setup your Python environment ASAP
- First recitation is Mon 1/21
 - Get help with your Python env.
 - Learn basics of arrays, plotting, etc
- HWo due NEXT WEEK (Wed 1/23 at 11:59pm)